Abstract

Machine learning is a concerned with the design and development of algorithms. Machine learning is a programming approach to achieve optimization. Classification is the prediction approach in data mining techniques. Decision tree algorithm is the most common classifier to build tree because of it is easier to implement and understand. Attribute selection is a concept by which to select more significant attributes in the given datasets. This Paper proposed a novel hybrid approach with a combination of rough set and Random Forest algorithm called Rough Set based Random Forest Classifier (RSRF Classifier) which is used to deal with uncertainties, vagueness, and ambiguity associated with datasets. In this approach, the selection of significant attributes based on rough set theory as an input to Random Forest classifier for constructing the decision tree which is more efficient and scalable approach as compare to related work for lymph disease diagnosis studies.
1. A. Verikas, A. Gelzinis, and M. Bacauskiene “Mining data with random forests: A survey and results of new tests” ELSEVIER, 2011
2. Ahmad Taher Azara, Hanaa Ismail Elshazlyb, Aboul Ella Hassanienb, and Abeer Mohamed Elkorany “A random forest classifier for lymph diseases” ELSEVIER, 2014
3. Qiang Shen and Richard Jensen “Rough Sets, their Extensions and Applications” IJAC, 2008
5. Joaquin Abellan, and Andres R. Masegosa “Bagging schemes on the presence of class noise in classification” ELSEVIER, 2012
6. Iffat A. Gheyas, and Leslie S. Smith “Feature subset selection in large dimensionality domains” ELSEVIER, 2010
7. Mohammad Lutfi Othman, Ishak Aris, Mohammad Ridzal Othman, and Harussaleh Osman “Rough-Set-and-Genetic-Algorithm based data mining and Rule Quality Measure to hypothesize distance protective relay operation characteristics from relay event report” ELSEVIER, 2011
8. Yiyu Yao “Rough-set concept analysis: Interpreting RS-definable concepts based on ideas from formal concept analysis” ELSEVIER, 2016

Index Terms
Computer Science Algorithms

Keywords
Machine learning, Rough Set, Decision Tree, Random Forest Classifier, Lymph disease